Efficiency of Simulated Annealing for Peptides with Increasing Geometrical Restrictions

CANAN BAYSAL, HAGAI MEIROVITCH

Supercomputer Computations Research Institute, Florida State University, Tallahassee, Florida 32306-4130

Received 4 February 1999; accepted 6 July 1999

ABSTRACT: Simulated annealing (SA) is a popular global minimizer that can conveniently be applied to complex macromolecular systems. Thus, a molecular dynamics or a Monte Carlo simulation starts at high temperature, which is decreased gradually, and the system is expected to reach the low-energy region on the potential energy surface of the molecule. However, in many cases this process is not efficient. Alternatively, the low-energy region can be reached more effectively by minimizing the energy of selected molecular structures generated along the simulation pathway. The efficiency of SA to locate energy-minimized structures within 5 kcal/mol above the global energy minimum is studied as applied to three peptide models with increasing geometrical restrictions: (1) The linear pentapeptide Leu-enkephalin described by the ECEPP potential, (2) a cyclic hexapeptide described by the GROMOS force field energy E_{GRO} alone, and (3) the same cyclic peptide with E_{GRO} combined with a restraining potential based on 31 proton–proton restraints obtained from nuclear magnetic resonance (NMR) experiments. The efficiency of SA is compared to that of the Monte Carlo minimization (MCM) method of Li and Scheraga, and to our local torsional deformations (LTD) method for the conformational search of cyclic molecules. The results for the linear peptide show that SA provides a relatively weak guidance towards the most stable energy region; as expected, this guidance increases for the cyclic peptide and the cyclic peptide with NMR restraints. However, in general, MCM and LTD are significantly more efficient than SA as generators of low-energy minimized structures. This suggests that

Correspondence to: H. Meirovitch; e-mail: hagai@scri.fsu.edu

Contract/grant sponsor: SCRI; contract/grant number DE-FC05-85ER250000.

Contract/grant sponsor: DOE; contract/grant number: DE-F605-95ER62070.

CCC 0192-8651 / 99 / 151659-12

LTD might provide a better search tool than SA in structure determination of protein regions for which a relatively small number of restraints are provided by NMR. © 1999 John Wiley & Sons, Inc. J Comput Chem 20: 1659–1670, 1999

Keywords: simulated annealing; molecular dynamics; Monte Carlo; peptides; NMR distance restraints; energy minimization

Introduction

Predicting the 3D native structure of a protein from its sequence of amino acid residues is an extremely difficult problem in global optimization. The energy surface of a protein, which is commonly described by an empirical potential function (force field),[1–3] consists of a tremendous number of local energy wells (the corresponding regions in conformational space have been called in our previous studies localized microstates); hence, locating the global energy minimum (GEM) resembles the search for a needle in a haystack. Furthermore, because the free energy rather than the energy is the correct criterion of stability, entropic effects should be taken into account as well.[4] Therefore, in principle, a suitable strategy would be to first identify the GEM structure and a set of low-energy minimized structures within some energy range above the GEM, assuming that the structure with the localized microstate of global free energy minimum resides within this range. Then, the free energy of the corresponding localized microstates can be calculated at least approximately (e.g., using the harmonic approximation).[5–8]

However, this picture is still somewhat idealized because molecular dynamics (MD) simulations have shown that the molecule will visit a localized microstate only for a very short time.[9,10] On the other hand, it will stay for a much longer time within a larger potential energy well, defined over a wide microstate (e.g., the region in conformational space covered by the protein fluctuations around its native structure); a wide microstate typically includes many localized ones. Thus, one should find the most stable wide microstate, i.e. that with the minimum free energy, which has more experimental significance than the most stable localized microstate.

The strategy described above is not feasible for a large protein because of the difficulty to find its GEM and other low-energy structures, and hence, the most stable wide microstate. However, this strategy can be used efficiently for determining the structure of a short protein segment such as a missing surface loop in homology studies. This approach would also be applicable to a short peptide, which, under certain solvent conditions, might reside in a single wide microstate or populate several wide microstates in thermodynamic equilibrium. The latter situation is frequently reflected in nuclear magnetic resonance (NMR) data, where the observed nuclear Overhauser effect (NOE) intensities are only compatible with a molecule that interconverts between several different structures.[11–16]

Analysis of such NOE data requires identifying the most stable wide microstates and calculating their populations.[12,17] With a new methodology (see refs. 7, 18, and 19), this is achieved by first carrying out an extensive conformational search for the GEM structure and the energy minimized structures within 2–3 kcal/mol above it; these structures are expected to reside within the most stable wide microstates. From this group one selects a smaller set of structures that are significantly different, which represents the most stable wide microstates. Each structure becomes a seed for an MC or MD simulation, which spans the corresponding wide microstate; the free energy is calculated with the local states method, from which the relative populations are obtained.[20–23]

Note that spanning the wide microstate by the simulation is, to a large extent, independent of the seed if the latter is of relatively low energy. Still, the search should be extensive but not necessarily exhaustive; i.e., it is sufficient to represent each wide microstate by only one low energy seed. On the other hand, in the organic chemistry community, the common approach has been to carry out an exhaustive search for all the energy minimized structures or for those within relatively large energy ranges above the GEM.[24–31]

This discussion already demonstrates the need to develop efficient procedures for global energy minimization, which provide the GEM as well as a group of other low-energy minimized structures. Applying the above-mentioned methodology to

cyclic peptides, we have used a new efficient search technique, the local torsional deformation (LTD) method, which is also suitable to handle protein segments (loops) in a dense protein environment.[32–34] LTD is a step-by-step procedure designed for a force field based on flexible geometry (i.e., flexible bond lengths and angles). Thus, at every step, several significant rotations around torsional angles are applied along the chain, each affecting only a small group of neighboring atoms. The deformed structure is energy minimized, and this minimized structure is accepted or rejected according to a selection criterion, which directs the system towards the low-energy regions; we use the Monte Carlo minimization (MCM) method of Li and Scheraga.[35] Thus LTD(MCM) is related to the group of techniques developed in the organic chemistry community for small molecules mentioned earlier,[24–31] which are based on energy minimization.

Simulated annealing (SA)[36–39] is a popular global optimizer based on an MD or a usual Metropolis MC simulation that starts at a high temperature, T, where T is decreased gradually; the system is expected to be driven towards the lowest energy region. SA can be applied relatively easily to complex systems, in particular to cyclic peptides, and it has the advantage that no energy minimization is required. However, in practice, when used with a regular force field, an SA process does not always lead the system to the most stable region at the end of the simulation. On the other hand, very low energies have been obtained by minimizing the energy of structures generated at the initial and intermediate stages of an SA process.[40] Thus, applying the latter strategy can make SA a convenient tool for locating the GEM and other low-energy minimized structures.

Thus, in a previous article, the efficiency of SA(MC) was compared to that of MCM, as applied to the linear pentapeptides Met- and Leu-enkephalin described by the ECEPP potential.[41] MCM was found to be significantly more efficient than SA; i.e., for a flexible linear peptide, which is typically characterized by a relatively large and complex energy landscape, the SA process does not provide strong enough guidance towards the low-energy region. This suggests that the performance of SA might improve for a cyclic peptide, which is much more constrained in space than a linear peptide; further improvement is expected in the structure determination of peptides or proteins based on proton–proton distances obtained from NMR (NOEs). In this case, one optimizes a function that consists of the force-field energy and a distance-restraining potential;[42–49] therefore, the available conformational space of the molecule is decreased significantly.

In the present work we apply both LTD(MCM) and SA(MD) to the cyclic hexapeptide cyclo-(D-Pro^1-Phe^2-Ala^3-Ser^4-Phe^5-Phe^6), which was studied by Kessler et al. using NMR.[50] They obtained 31 NOEs and several ^3J coupling constants leading to the conclusion that the molecule significantly populates at least two wide microstates. Structural analysis based on these data has also been carried out recently with our methodology;[34,51] as in these studies, the peptide is modeled here with the GROMOS force field alone, and together with a restraining potential based on the 31 NOEs. We compare the efficiency of SA and LTD to generate the GEM structure and other energy minimized structures within 5 kcal/mol of the GEM; the criterion for variance of two structures is that at least one dihedral angle differs by $d = 2°$ or more. Also, two conformations are defined as significantly different, i.e., they are expected to reside in different wide microstates, if at least one dihedral angle differs by $d = 40°$ or more; in a recent work[51] this criterion has been found more suitable than $d = 60°$ used previously. Correspondingly, we also check the efficiency of the two methods to generate significantly different structures, i.e., different wide microstates. To broaden the scope of this study, we also analyze with the $d = 40°$ criterion a set of energy minimized structures of Leu-enkephalin obtained in ref. 41 using the ECEPP potential. Finally, the ability of the methods to generate structures with a large number of satisfied NOEs is investigated.

Methods

THE MODEL

The intramolecular interactions of cyclo-(D-Pro^1-Phe^2-Ala^3-Ser^4-Phe^5-Phe^6) are described by the GROMOS 37D4 united atom force field, which define the molecular energy E_{GRO}.[3] These interactions include harmonic bond stretching and bond angle-bending potentials, proper and improper torsional potentials, and nonbonded electrostatic and 6–12 Lennard-Jones (LJ) interactions.

$$E_{\mathrm{GRO}} = \sum_{n=1}^{N_b} \frac{1}{2} K_{b_n} [\, b_n - b_{0_n}]^2$$

$$+ \sum_{n=1}^{N_\theta} \frac{1}{2} K_{\theta_n} [\theta_n - \theta_{0_n}]^2 + \sum_{n=1}^{N_\xi} \frac{1}{2} K_{\xi_n} [\xi_n - \xi_{0_n}]^2 + \sum_{n=1}^{N_\phi} K_{\phi_n} [1 + \cos(m_n \phi_n - \delta_n)] + \sum_{i,j;\, j \geq i+4} \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + \frac{C_{ij}}{r_{ij}} \qquad (1)$$

In eq. (1), b_n is the bond length, θ_n, ϕ_n, and ξ_n are the valence, dihedral, and improper dihedral angles, respectively; the subscript 0 denotes equilibrium values. N_b, N_θ, N_ϕ, and N_ξ are the number of bonds and valence, dihedral, and improper dihedral angles in the molecule, respectively; K_{b_n}, K_{θ_n}, K_{ϕ_n}, and K_{ξ_n}, are the corresponding force constants, m_n is an integer within the interval [1, 6], and $\delta_n = 0$ or π. A_{ij} and B_{ij} ar Lennard-Jones force constants for atom pair i and j separated by distance r_{ij}. C_{ij} is the corresponding electrostatic interaction constant that is defined for the dielectric constant $\epsilon = 1$. The nonbonded interactions were not truncated. The hydrogen atoms are treated as collapsed on their first neighboring atoms except for hydrogens bonded to a nitrogen or the oxygen of Ser. As a result, the cyclohexapeptide consists of 57 explicit atoms.

The above hexapeptide is studied by both LTD and SA using either E_{GRO} [eq. (1)], or E_{res}, where E_{res} also includes an unphysical restraining potential that forces the molecule to stay within the regions that satisfy the 31 experimental NOEs.[52] Denoting the upper and lower bound distances of proton pair m by d_m^u and d_m^l, respectively,[50] E_{res} is

$$E_{\mathrm{res}} = E_{\mathrm{GRO}} + \sum_{m=1}^{31} \left[\frac{1}{2} k_m (r_m - d_m^u)^2 + \frac{1}{2} k'_m (r_m - d_m^l)^2 \right] \qquad (2)$$

where r_m is the instantaneous distance of pair m; k_m and k'_m are given by

$$k_m = 0 \quad \text{for } r_m < d_m^u,$$
$$k'_m = 0 \quad \text{for } r_m > d_m^l,$$
$$k_m = k'_m = k_r \quad \text{otherwise}$$

where k_r is a force constant.

We also present results and analyze data for Leu-enkephalin (H-Tyr-Gly-Gly-Phe-Leu-OH) obtained in refs. 8 and 41. This molecule is modeled by ECEPP/2,[1,2] which includes nonbonded, electrostatic, torsional, and hydrogen-bond potentials, and is based on rigid geometry, i.e., constant bond lengths and angles; the peptide dihedral angles ω are fixed at 180°, and all the hydrogen atoms are described explicitly. Thus, a conformation of Leu-enkephalin is defined by the 10 backbone dihedral angles ϕ and ψ and the nine side-chain dihedral angles χ.

SIMULATED ANNEALING

Each SA(MD) run starts from a randomly chosen conformation, and the system is first equilibrated for 5 ps at a constant initial temperature, T_0. This temperature is then decreased by rescaling the atoms' velocities with the standard procedure proposed by Berendsen et al.[53] and adapted in GROMOS. The annealing process from T_0 to the final temperature T_f is carried out in n steps each of 0.5 fs, which is also the time step of the integration of the equations of motion. At every $n/5000$ steps, the coordinates of the current structure are saved in a file for a later analysis, which includes energy minimization carried out with the limited memory L-BFGS method;[54] thus, 5000 structures are retained from each run. Because the experimental evidence suggests that all the peptide bonds are in the *trans* position, a restraining potential is applied to them to prevent a *trans-cis* barrier crossing at high temperatures. This potential is removed for steps j, where $T_j < 500$ K, and is not applied in the minimization stage. The efficiency of SA is known to be system dependent; therefore, we test two cooling procedures, each for three different durations n, as described below. For each of the six cases, five independent runs starting from different initial structures are carried out; thus, a set of 25,000 structures are collected for each case. All the calculations of this work were performed on a Silicon Graphics O2 workstation.

With the exponential cooling procedure, the temperature at step j, T_j, is defined by

$$T_j = T_{j-1} \left(\frac{T_f}{T_0} \right)^{1/n} \qquad (3)$$

whereas in linear cooling, T_j is given by

$$T_j = T_{j-1} - \frac{T_0 - T_f}{n} \qquad (4)$$

To test the effect of the cooling duration on the efficiency, runs of $t = 50$ ps ($n = 10^5$), 200 ps ($n = 4 \times 10^5$), and 2 ns ($n = 4 \times 10^6$) are carried out. The six different cases are described in Table I.

THE LTD METHOD

The LTD process starts from an arbitrary, energy-minimized structure of the cyclic peptide, i with energy E_i^{m} (m stands for ''minimized''). i is deformed by applying several significant torsional rotations along the chain, where each rotation is localized, i.e., disrupts a bond neighbor to the rotated one (see details in Fig. 1). The energy is then minimized, leading to a trial conformation j (with energy E_j^{m}), which is accepted or rejected according to a selection procedure; the accepted conformation is then deformed and the process is repeated many times; obviously, this conformational change enables the molecule to rapidly cross energy barriers. The role of the selection procedure is to direct the search towards the low energy regions of conformational space, in particular towards the GEM. As a selection procedure, we use the Monte Carlo minimization (MCM) method of Li and Scheraga,[35] where j is accepted with a Metropolis MC probability p_{ij},

$$p_{ij} = \min\left\{1, \exp\left[-(E_j^{\mathrm{m}} - E_i^{\mathrm{m}})/k_{\mathrm{B}}T^*\right]\right\}. \quad (5)$$

T^* is a temperature efficiency parameter, which in our case is varied between 200 and 800 K in increments of 200 K every 200 MCM steps. Also, if the system becomes trapped in a particular structure, T^* is raised to 2000 K for 20 steps and then returned to its previous value. Because the experiment suggests that all the peptide bonds are in the *trans* position, the rarely generated *cis* conformations are discarded. As in our previous studies, the coordinates and energies of all the energy minimized structures, including those that have not been accepted by the Metropolis criterion, are stored in a file for further analysis.

It should be pointed out that the known applications of MCM to linear peptides, including that of ref. 41, are all based on nonlocalized dihedral angle changes, i.e., the whole chain segment following the rotated angle is moved; correspondingly, nonlocalized rotations are applied here to the side chain angles χ of the hexapeptide. The details of the LTD(MCM) procedure employed here are described in the ''Conformational Search'' section of ref. 34.

To base the comparison of SA and LTD on equal footing, we apply these methods for equal amounts of CPU time. We first determine the most efficient SA procedure, as described in the next section. Then, for this procedure, we find the total CPU time required by the five MD runs and the minimization of the 25,000 saved structures. We carry out five LTD runs, starting from different initial structures, such that altogether they consume the total SA CPU time. The energy minimized structures obtained from the five LTD runs are lumped together for further analysis. For simplicity, in what follows we denote LTD(MCM) by LTD.

TABLE I.
Summary of Different SA Simulations.

Cooling Type[a]	Length (ps)[b]	CPU time (s)[c]	E^d_{lowest}	No. of Unique Structures[e]
Exponential	50	43,950	13.100	2359
Exponential	200	45,378	11.349	7278
Exponential	2000	91,625	10.341	13,626
Linear	50	46,352	12.225	3242
Linear	200	47,172	10.332	9724
Linear	2000	96,735	10.335	18,103

[a] Exponential and linear cooling are defined by eqs. (3) and (4), respectively.
[b] Simulation time of each MD run; 5000 structures with equal time spacing are saved from each run; a set of five such runs are carried out for each case.
[c] Total CPU time of the five MD runs and the energy minimization of the 25,000 saved structures.
[d] Lowest minimized energy (in kcal / mol) obtained in each case; the GEM is 10.328 kcal / mol.
[e] Criterion of uniqueness is that at least one dihedral angle differs by $d \geq 2°$.

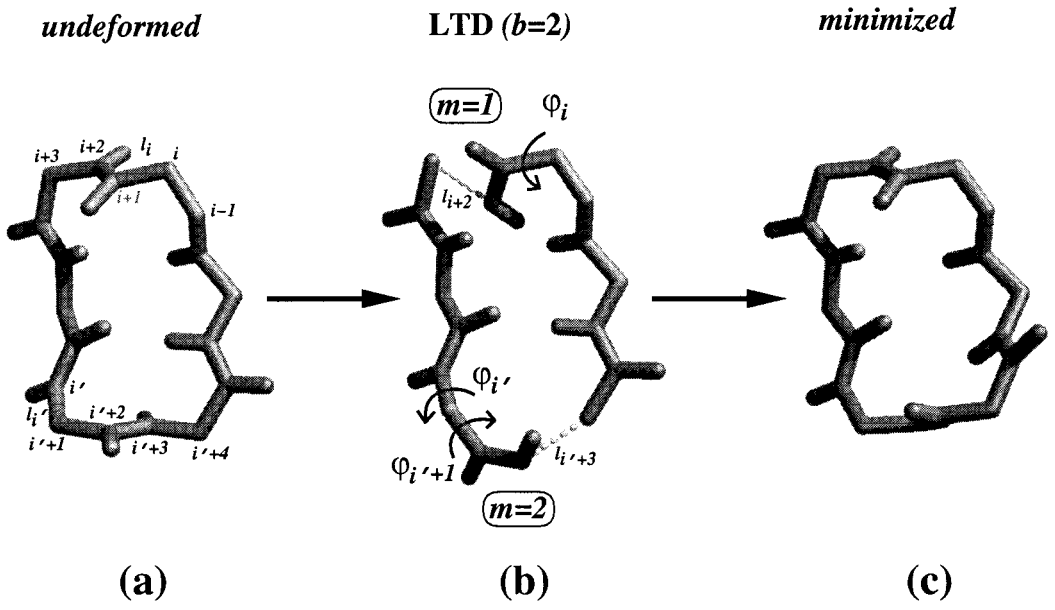

FIGURE 1. Local rotations of LTD illustrated for the cyclo-$(Gly)_6$ molecule. (a) The undeformed structure. (b) Two simultaneous local rotations. A single local rotation is applied to the upper part of the molecule by changing ϕ_i around the single bond l_i; this affects the backbone atom at location $i + 2$ as well as the side chain atoms attached to the atoms at $i + 1$ and $i + 2$. The affected bond l_{i+2} is denoted by a dashed line. Two successive rotations at the lower part are performed around bonds $l_{i'}$ and $l_{i'+1}$ and the affected bond is $l_{i'+3}$; in this case the positions of atoms $i' + 2$, $i' + 3$ and the side chains attached to atoms $i' + 1$, $i' + 2$, and $i' + 3$ are changed. (c) The energy minimized structure that differs from the initial structure around the deformed regions.

Results and Discussion

OPTIMIZATION OF SA

An SA run is controlled by T_0, T_f, the simulation length, and the type of cooling procedure used (in our case exponential or linear). T_0 should be high enough to ensure a randomized sampling of the conformational space in the early stage of the process; T_f, on the other hand, should be low enough to keep the molecule in the final wide microstate attained. The values used here, $T_0 =$ 1200 K and $T_f = 200$ K satisfy these requirements, as is demonstrated in Figure 2 for a run based on slow exponential cooling. Thus, while the energy shows a clear trend to decrease along the whole annealing pathway (Fig. 2a), the minimized energy (Fig. 2b) fluctuates strongly between $\sim$10–35 kcal/mol for $T > \sim 500$ K, but very little for $T < \sim 400$ K; this means that for the latter range of temperatures the molecule has already settled in some wide microstate. Similar results have been obtained for the other runs.

Previous SA studies[40] have shown that very low-energy minimized structures can already be obtained in the early and middle stages of the simulation, while sometimes the final minimized energy is higher. In Figure 2 very low-energy minimized structures also appear in the early stage of the SA run. This suggests that the correct strategy would be to select structures along the cooling pathway rather than only from its final stage, as has been discussed in the Introduction.

Next, we have sought to find the most efficient cooling procedure and optimize the simulation length for the values of T_0 and T_f used. In Table I, results are presented for the total CPU time required for the five SA simulations and the energy minimization of the 25,000 saved structures using E_{GRO} [eq. (1)]. The table also shows the lowest energy attained and the number of different structures (based on the $d = 2°$ criterion) found in each case. Evidently, linear cooling is the superior procedure, because it has found the lowest energy minima and has generated a larger number of different energy minimized structures than exponential cooling for any given simulation length. Also, fast cooling is clearly less efficient than moderate and slow cooling, where the latter two procedures are found to be comparable in terms of the number of unique structures generated per unit

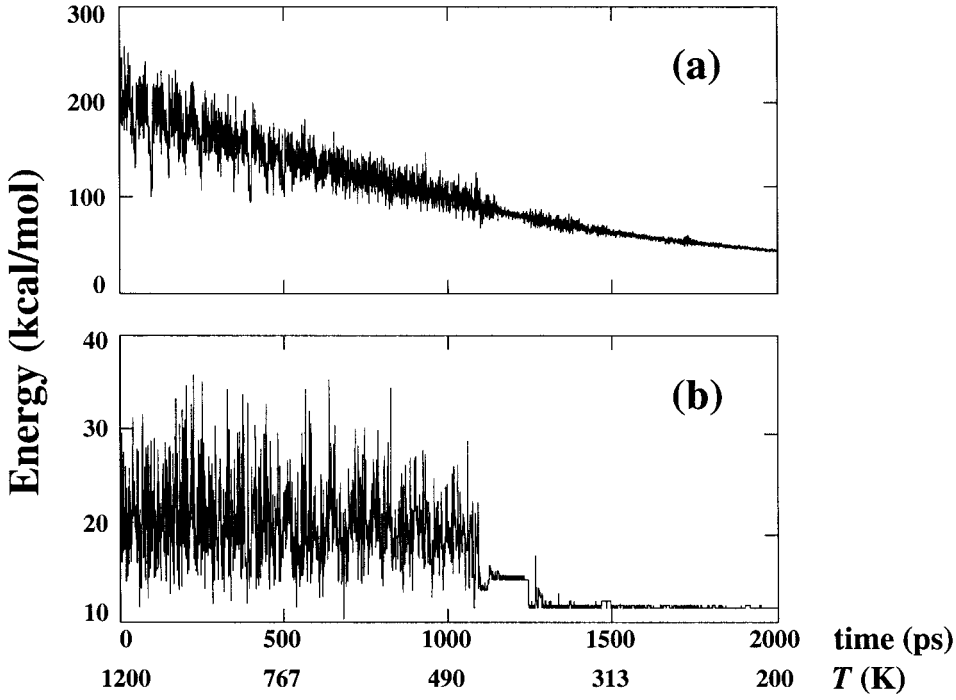

FIGURE 2. Variation of (a) the energy and (b) the minimized energy, in the course of an SA run with exponential cooling plotted as a function of time (ps) and system temperature (K).

time. However, with slow cooling the lowest energy region (i.e., $E < 10.4$ kcal/mol) is accessed in three of the five runs, whereas with moderate cooling this region is reached in only a single run. Therefore, the slow linear cooling has been accepted by us as the most efficient SA procedure to be compared to LTD. It should be pointed out that the average CPU time required for the minimization of a structure obtained in the fast and medium cooling is larger than that needed for the slow cooling, because the energies of unminimized structures (e.g., see Fig. 2a) are higher for the former.

COMPARISON OF SA AND LTD WITH E_{GRO}

Using the $d = 40°$ criterion, we have first analyzed the energy minimized structures obtained by MCM and SA in previous work[8,41] for the linear pentapeptide Leu-enkephalin described by ECEPP. The results are presented in Table II, which also includes the $d = 2°$ results derived in ref. 41 for comparison. Before discussing these results, it should be pointed out again that the $d = 40°$ criterion is used to provide seeds for MC simulations that span the corresponding wide microstates. These samples cover relatively large regions of conformational space and, therefore, their distinctiveness should be verified (for details, see refs. 18 and 19). However, for a long peptide this angular criterion might prove insufficient because small angular changes along the backbone of a given structure can result in a significantly different con-

TABLE II.
Number of Energy Minimized Structures within Energy Bins of Size 0.5 kcal / mol above the GEM Obtained by SA and MCM for Leu-Enkephalin Described by the ECEPP Potential.[a]

	$d = 2°$[b]		$d = 40°$	
Bin (kcal / mol)	SA	MCM	SA	MCM
GEM[c]	1 / 15	12 / 14		
0.0 –0.5	3	8	3	8
0.5 –1.0	6	12	6	10
1.0 –1.5	7	20	7	16
1.5 –2.0	8	75	8	52
2.0 –2.5	10	118	9	100
2.5 –3.0	14	258	13	208
3.0 –3.5	22	384	16	308
3.5 –4.0	40	614	34	495

[a] For simulation details see refs. 8 and 41.
[b] Results from ref. 41.
[c] (The number of runs that found GEM) / (total number of runs).

formation, which might be interpreted mistakenly as pertaining to the wide microstate of the original structure. Therefore, in principle, one should also use, in addition to the angular criterion, a criterion based on the rms deviation between structures. However, for the short chains studied here the angular criterion is sufficient. Thus, we checked, for example all the structures of the pentapeptide with minimized energy within 10 kcal/mol above the GEM and dihedral angles that do not deviate by more than 40° from those of the GEM structure. We have found 261 such structures which all have a 2–5 βII′ turn, meaning that they pertain to the same wide microstate. A similar analysis for the cyclic hexapeptide leads to the same conclusion. For example, the structure with the largest deviation of backbone dihedral angles from the GEM angles (which, however, do not exceed 40°) has a backbone rms deviation from the GEM structure of only 0.5 Å.

Table II shows that for both criteria, MCM is significantly more efficient than SA, i.e., the number of structures generated by SA in all the bins is much smaller than that of MCM, while the ratio of the required CPU times $t(\mathrm{SA})/t(\mathrm{MCM}) \approx 15$ is large. As has been pointed out previously, this reflects the fact that the force field alone provides only a weak guidance towards the low energy regions of a linear flexible molecule.

The same analysis has been applied to the 25,000 structures generated in five SA(E_{GRO}) runs and then energy minimized, and to the 43,155 energy minimized structures obtained in five LTD runs that started from the SA initial structures. To recall, the LTD and SA calculations each has taken the same amount of CPU time, $t = 96{,}735$ s. The results are presented in Table III, which shows that for the $d = 2°$ criterion the number of structures found for the hexapeptide is significantly larger than that found for the linear pentapeptide in Table II; this is attributed to the rigid geometry vs. flexible geometry force fields used for these peptides, respectively. With LTD, the lowest energy structure (10.328 kcal/mol) was reached in all five runs. This structure was generated several times in each run, and on average, appeared for the first time after 1378 ± 293 minimizations, requiring ~ 600 s of CPU time. This gives reason to believe that this is the GEM structure.

Unlike LTD, with SA, structures with energy smaller than 10.4 kcal/mol were found in only three of the five SA runs; these structures are very similar to the GEM structure, but have slightly higher energy values. LTD is also significantly more efficient than SA for generating energy minimized conformations differing by $d = 2°$; this applies to all of the bins besides the first bin of 0.5 kcal/mol, where the numbers are comparable and thus constitute an improvement over the corresponding results in Table II.

TABLE III.
Number of Energy Minimized Structures within Energy Bins of Size 0.5 kcal / mol above the GEM Obtained by SA and LTD for the Cyclic Peptide Using E_{GRO}.

	$d = 2°$		$d = 40°$	
Bin (kcal / mol)	SA	LTD	SA	LTD
GEM[a]	3 / 5[b]	5 / 5[c]		
0.0 –0.5	6	7	2	2
0.5 –1.0	5	58	2	3
1.0 –1.5	28	584	7	7
1.5 –2.0	30	691	6	7
2.0 –2.5	87	905	11	17
2.5 –3.0	59	710	20	25
3.0 –3.5	239	1186	36	51
3.5 –4.0	275	1202	54	78
4.0 –4.5	539	1526	81	102
4.5 –5.0	520	1846	98	135

[a] (The number of runs with structures of $E < 10.4$ kcal / mol) / (total number of runs); GEM is 10.328 kcal / mol.
[b] The GEM is not found in any of the five runs initiated from randomly selected starting structures; lowest energy found is 10.335 kcal / mol.
[c] The GEM is found in all five runs; it is reached after an average of 1378 ± 293 LTD minimizations.

All these results demonstrate that for the cyclic molecule as well as the linear one, the MCM selection procedure strongly drives the system towards the low-energy region, while SA provides a much more random sampling, i.e., it efficiently generates the high-energy structures that constitute the majority. Indeed, we have found (data not shown) that SA becomes more efficient than LTD for locating structures with energy greater than 10 kcal/mol above the GEM, which are, however, thermodynamically insignificant at 300 K.

The analysis based on $d = 40°$ shows that the number of wide microstates is significantly smaller than that of the localized ones. To further verify this, a larger set of $N = 82{,}868$ structures was generated by both SA and LTD and was analyzed with $d = 40°$. Indeed, the number of different minima in the first five bins remained equal to that found by LTD, whereas for the sixth bin it only slightly increased from 25 to 28. This relatively small number of wide microstates stems from the

fact that the low-energy structures are compact, i.e., the side chains collapse on the cyclic backbone, and large rotations of the χ angles generally destroy this compactness leading to high-energy structures. For the linear Leu-enkephalin, on the other hand, the side chain–side chain and side chain-backbone interactions are weaker and, therefore, significant rotations of the χ angles still lead to low energy structures for basically the same backbone conformation; thus, the number of the wide microstates in Table II is relatively high.

Table III also reveals that the results of the two methods in the first four bins are comparable, which demonstrates an improvement in the efficiency of SA compared to the corresponding results in Table II; notice that the improvement in the first two bins can partially be attributed to the small number of wide microstates. For the higher energy bins, LTD always produces more structures than SA; however, the difference in efficiency between the two methods is not as significant as that observed in Table II and for $d = 2°$ in Table III. This leads to the following picture: with SA, the small number of low-energy wide microstates are visited (not necessarily in the final stage of the simulation, see Fig. 1) only for relatively short times, which allows locating a small number of energy minimized structures differing by $d = 2°$. For example, the seven wide microstates of the third bin were all found by SA, which however, could find only 28 of the 584 energy minimized structures located by LTD.

In summary, for both peptides, LTD (or MCM) is significantly more efficient than SA; however, as expected, the relative efficiency of SA improves in going from the linear to the cyclic peptide, as demonstrated by the results for the first bin ($d = 2°$), and those obtained with $d = 40°$. Obviously, this improved efficiency will decrease with increasing the size of the cycle peptide.

COMPARISON OF SA AND LTD WITH E_{res}

The improvement of SA for the cyclic peptide is due to the stronger geometrical constraints imposed on the cyclic molecule. Therefore, one would expect the relative efficiency of SA to increase further, by adding to the force field energy a potential function based on distance and angular restraints, as is commonly done in structure determination of proteins based on multidimensional NMR data. To check this, we carried out SA and LTD runs with E_{res} [eq. (2)] using restraining constants $k_r = 2.39$ and 4.78 kcal mol^{-1} Å^{-2} (corresponding to 1000 and 2000 kJ mol^{-1} nm^{-2}, which are typical values used for such peptide calculations). The numbers of different energy minimized structures obtained are presented in Table IV for $d = 2°$ and $d = 40°$. It should be pointed out that the restraining potential is expected to further complicate and ''enrich'' the energy surface of the molecule with energy minima. Indeed, the number of structures obtained with $d = 2°$ for both k_r

TABLE IV.
Number of Energy-Minimized Structures within Energy Bins of Size 0.5 kcal / mol above the GEM Obtained by SA and LTD for the Cyclic Peptide Using E_{res}.[a]

	$d = 2°$				$d = 40°$			
	$k_r = 2.39$[b]		$k_r = 4.78$[c]		$k_r = 2.39$[b]		$k_r = 4.78$[c]	
Bin (kcal / mol)	SA	LTD	SA	LTD	SA	LTD	SA	LTD
0.0 –0.5	7	113	21	199	2	2	2	1
0.5 –1.0	17	208	56	388	2	3	6	6
1.0 –1.5	26	397	107	599	9	11	13	5
1.5 –2.0	141	642	130	593	13	16	21	15
2.0 –2.5	203	989	314	1158	21	29	48	25
2.5 –3.0	157	1202	334	1261	43	50	42	35
3.0 –3.5	194	1300	762	1726	52	81	110	68
3.5 –4.0	432	1516	598	1813	96	99	112	63
4.0 –4.5	524	1539	686	2135	137	135	143	76
4.5 –5.0	786	1844	965	2309	184	188	200	105

[a] k_r is the force constant of the distance restraining potential in eq. (2).
[b] The GEM is 19.794 kcal / mol.
[c] The GEM is 27.431 kcal / mol.

values is larger in most cases in Table IV than in Table III.

Table IV reveals that, in accord with the previous tables, for $d = 2°$ and both values of k_r, LTD is significantly more efficient than SA in finding low energy-minimized structures. On the other hand, for $d = 40°$, the SA results for $k_r = 2.39$ kcal mol^{-1} Å^{-2} become approximately comparable to those of LTD, and for $k_r = 4.78$ kcal mol^{-1} Å^{-2}, the number of structures located by SA is larger in most cases than that of LTD. This trend is in accord with the expectation stated above that increasing the restraining potential would increase the efficiency of SA for finding the wide microstates. However, it should be pointed out that part of this improvement of SA over LTD might be due to the reduced efficiency of the MCM procedure as the restraining potential is increased; thus, the energy differences between the wide microstates increase and the corresponding transition probabilities decrease. In other words, the system will stay for long simulation times in a relatively small number of wide microstates which, therefore, will be sampled efficiently; this is reflected by the larger number of structures found by LTD than SA for $d = 2°$ in Table IV. The increase of the restraining potential is, therefore, equivalent to freezing the molecule environment, i.e., lowering T^* in eq. (5). To check this point further, we carried out another conformational search with LTD for $k_r = 4.78$ kcal mol^{-1} Å^{-2}, where T^* is varied between 600–1200 K instead of 200–800 K. In this case, LTD becomes equally efficient to SA (data not shown), implicating that T^* should be optimized according to the potential function used.

Another aspect of structure determination is related to the efficiency of a method to locate structures that satisfy most of the NOEs. Therefore, we have plotted the number of unique structures that satisfy a given number of NOEs for SA and LTD. Figure 3, which displays such a distribution for $k_r = 4.78$ kcal mol^{-1} Å^{-2} with $d = 2°$, constitutes a typical example. The figure shows that SA is more efficient than LTD in finding structures with a large number of satisfied NOEs (> 21). This, however, is not surprising, and in fact, demonstrates a weakness of SA, i.e., its inefficiency to generate the low energy structures. Thus, the low minimized energy is a result of an optimal balance between E_{GRO} and the restraining potential, which typically occurs for a relatively small number of satisfied NOEs. On the other hand, the high minimized energy is a result of a relatively low restraining potential (i.e., a large number of NOEs are satisfied) at the expense of an increase in E_{GRO} due to bad overlaps between atoms. Because SA visits the high-energy region more effectively than LTD, it is better in locating the structures with a large number of satisfied NOEs. Obviously, one would expect the force field energy to be minimal for the experimental structure; the deviation of E_{GRO} from this behavior stems from the fact that it does not take into account solvent effects.[34]

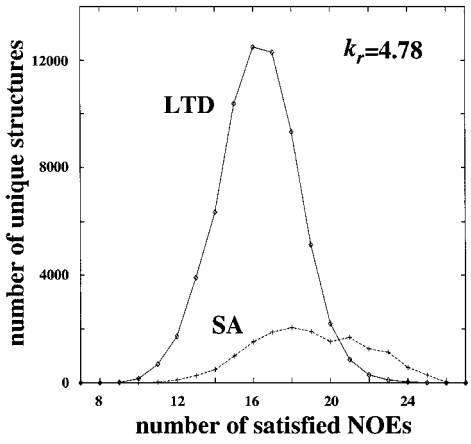

FIGURE 3. The number of unique structures (with $d = 2°$) for $k_r = 4.78$ kcal mol^{-1} Å^{-2} vs. the number of satisfied NOEs, found during LTD (solid line) and SA (dashed line) runs. Although the total number of unique structures found by SA is much smaller than LTD, the former outnumbers the latter in the region with more than 21 satisfied distances. Both methods satisfy a maximum of 26 distances out of the 31 experimentally determined ones.

To further illustrate this phenomenon, we define $\bar{I}$, the average number of NOEs found in a given conformational search,

$$\bar{I} = \frac{\sum_{i=1}^{31} i \cdot n_i}{\sum_{i=1}^{31} n_i} \tag{6}$$

where n_i is the number of unique structures possessing i NOEs. We have found that for $k_r = 0$, $\bar{I}_{LTD} = 14.0$ and $\bar{I}_{SA} = 15.1$. As expected, for $k_r = 2.39$ kcal mol^{-1} Å^{-2}, these numbers increase, becoming $\bar{I}_{LTD} = 15.0$ and $\bar{I}_{SA} = 17.1$ and for $k_r = 4.78$ they increase further to $\bar{I}_{LTD} = 16.2$ and $\bar{I}_{SA} = 18.9$. Also, the expectation $\bar{I}_{LTD} < \bar{I}_{SA}$ for each k_r

value is always born out, where $\bar{I}_{SA}$ increases faster than $\bar{I}_{LTD}$ for larger values of k_r.

Conclusions

We have pointed out in the Introduction that SA is a convenient global minimizer, in particular for cyclic peptides; we have also emphasized that its efficiency can be improved by minimizing the energy of structures along the simulation pathway. The main aim of this article has been to study the relative efficiency of SA for generating the GEM and other energy minimized structures as applied to different potential functions, which impose increasing geometrical constraints on the molecule. Thus, three cases were examined: the linear pentapeptide Leu-enkephalin described by ECEPP, a cyclic hexapeptide described by GROMOS, and by the GROMOS energy and a restraining potential based on NMR distances. The SA results were compared to the results obtained with LTD and MCM.

In all the cases studied, LTD has been found to be significantly more efficient than SA in locating the GEM and low energy minimized structures (< 5 kcal/mol of the GEM) that differ by the 2° criterion. On the other hand, the efficiency of SA for generating structures with a variance criterion of 40° (which are expected to reside in different wide microstates), has been found to improve as the energy function imposes stronger geometrical restraints on the molecule. Thus, if one is interested in finding all the energy minima within a given energy range (as is the case in the organic chemistry community), LTD and methods related to it are a much better choice than SA. Also, LTD appears to be the more suitable method for our methodology[7,18,19] because of its ability to search each wide microstate extensively, providing thereby a rich selection of low energy seeds for the MC or MD simulation. Furthermore, for larger cyclic peptides, LTD is expected to be a better tool than SA for identifying the wide microstates.

When enough NMR distance restrains are available, SA can be used successfully for structure determination of proteins based on a potential similar to that defined in equation 2.[42–45] However, the present study suggests that when the number of restraints is relatively small over the whole molecule or in a specific region, LTD might provide a better global optimizer that SA; this expectation will be checked in future work.

Acknowledgments

Some of the calculations were carried out by Lauren Harris and Mariah Jensen during the Research Science Mentorships Program at SCRI, organized by Leon County Academic Resource Center.

References

1. Momany, F. A.; McGuire, R. F.; Burgess, A. W.; Scheraga, H. A. J Phys Chem 1975, 79, 2361.
2. Sippl, M. J.; Némethy, G.; Scheraga, H. A. J Phys Chem 1984, 88, 6231.
3. van Gunsteren, W. F.; Berendsen, H. J. C. Groningen Molecular Simulation (GROMOS) Library Manual, Biomos: Nijenborgh 16 9747 AG Groningen NL, 1987.
4. Vásquez, M.; Némethy, G.; Scheraga, H. A. Chem Rev 1994, 94, 2183.
5. Gibson, K. D.; Scheraga, H. A. Physiol Chem Phys 1969, 1, 109.
6. Gō, N.; Scheraga, H. A. J Chem Phys 1969, 51, 4751.
7. Meirovitch, H.; Meirovitch, E.; Lee, J. J Phys Chem 1995, 99, 4847.
8. Meirovitch, H.; Meirovitch, E. J Comput Chem 1997, 18, 240.
9. Elber, R.; Karplus, M. Science 1987, 235, 318.
10. Stillinger, F. H.; Weber, T. A. Science 1984, 225, 983.
11. Shenderovich, M. D.; Nikiforovich, G. V.; Saulitis, J. B.; Chipens, G. Biophys Chem 1988, 31, 163.
12. Kessler, H.; Griesinger, C.; Lautz, J.; Müller, A.; van Gunsteren, W. F.; Berendsen, H. J. C. J Am Chem Soc 1988, 110, 3393.
13. Landis, C.; Allured, V. S. J Am Chem Soc 1991, 113, 9493.
14. Nikiforovich, G. V.; Prakash, O.; Gehrig, C. A.; Hruby, V. J. J Am Chem Soc 1993, 115, 3399.
15. Ulyanov, N. B.; Schmitz, U.; Kumar, A.; James, T. L. Biophys J 1995, 68, 13.
16. Bonvin, A. M. J. J.; Brünger, A. T. J Mol Biol 1995, 250, 80.
17. Brüschweiler, R.; Roux, B.; Blackledge, M.; Griesinger, C.; Karplus, M.; Ernst, R. R. J Biomol NMR 1991, 1, 3.
18. Meirovitch, E.; Meirovitch, H. Biopolymers 1996, 38, 69.
19. Meirovitch, H.; Meirovitch, E. J Phys Chem 1996, 100, 5123.
20. Meirovitch, H. Chem Phys Lett 1977, 45, 389.
21. Meirovitch, H. Phys Rev A 1985, 32, 3709.
22. Meirovitch, H.; Vásquez, M.; Scheraga, H. A. Biopolymers 1987, 26, 651.
23. Meirovitch, H.; Kitson, D. H.; Hagler, A. T. J Am Chem Soc 1992, 114, 5386.
24. Saunders, M. J Am Chem Soc 1987, 109, 3150.
25. Gotō, H.; Ōsawa, E. J Am Chem Soc 1989, 111, 8950.
26. Chang, G.; Guida, W. C.; Still, W. C. J Am Chem Soc 1989, 111, 4379.

27. Saunders, M.; Houk, K. N.; Wu, Y.-D.; Still, W. C.; Lipton, M.; Chang, G.; Guida, W. C. J Am Chem Soc 1990, 112, 1419.
28. Goodman, J. M.; Still, W. C. J Comput Chem 1991, 12, 1110.
29. Weinberg, N.; Wolfe, S. J Am Chem Soc 1994, 116, 9860.
30. Kolossváry, I.; Guida, W. C. J Am Chem Soc 1996, 118, 5011.
31. Wang, C.-S. J Comput Chem 1997, 18, 277.
32. Baysal, C.; Meirovitch, H. J Chem Phys 1996, 105, 7868.
33. Baysal, C.; Meirovitch, H. J Phys Chem A 1997, 101, 2185.
34. Baysal, C.; Meirovitch, H. J Am Chem Soc 1998, 120, 800.
35. Li, Z.; Scheraga, H. A. Proc Natl Acad Sci USA 1984, 84, 6611.
36. Kirkpatrick, S.; Gelatt, Jr., C. D.; Vecchi, M. P. Science 1983, 220, 671.
37. Wilson, S. R.; Cui, W. Biopolymers 1990, 29, 225.
38. Wilson, S. R.; Cui, W.; Moskowitz, J. W.; Schmidt, K. E. J Comput Chem 1991, 12, 342.
39. Okamoto, Y.; Kikuchi, T.; Kawai, H. Chem Lett 1992, 3, 1275.
40. Simmerling, C.; Elber, R. J Am Chem Soc 1994, 116, 2534.
41. Meirovitch, H.; Vásquez, M. J Mol Struct (THEOCHEM) 1997, 398–399, 517.
42. Nilges, M.; Clore, G. M.; Gronenborn, A. M. FEBS Lett 1988, 129.
43. Brünger, A. T.; Nilges, M. Quart Rev Biophys 1993, 26, 49.
44. Brünger, A. T.; Adams, P. D.; Rice, L. M. Structure 1997, 5, 325.
45. Nilges, M.; O'Donoghue, S. I. Prog Nucl Mag Res Spectrosc 1998, 107.
46. Coles, M.; Sowemimo, V.; Scanlon, D.; Munro, S. L. A.; Craik, D. J. J Med Chem 1993, 36, 2658.
47. Ma, S.; McGregor, M. J.; Cohen, F. E.; Pallai, P. V. Biopolymers 1994, 34, 987.
48. Pfeifer, M. E.; Linden, A.; Robinson, J. A. Helv Chim Acta 1997, 80, 1513.
49. Liwo, A.; Tempczyk, A.; Oldziej, S.; Shenderovich, M. D.; Hruby, V. J.; Talluri, S.; Ciarkowsky, J.; Kasprzykowski, F.; Lankiewicz, L.; Grzonka, Z. Helv Chim Acta 1997, 80, 1513.
50. Kessler, H.; Matter, H.; Gemmecker, G.; Kottenhahn, M.; Bats, J. W. J Am Chem Soc 1992, 114, 4805.
51. Baysal, C.; Meirovitch, H. Biopolymers 1999, 50, 329.
52. Zuiderweg, E. R. P.; Scheek, R. M.; Boelens, R.; van Gunsteren, W. F.; Kaptein, R. Biochimie 1985, 67, 707.
53. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Engberts, J. B. F. N. J Chem Phys 1984, 81, 3684.
54. Liu, D. C.; Nocedal, J. Technical Report NAM03, Department of Electrical Engineering and Computer Science, Northwestern University (1988).